AD-A110 926 ARMY ENGINEER TOPOGRAPHIC LABS FORT BELVOIR VA F/G 9/2
A NAMES INPUT STATION FOR AUTOMATED FOREIGN NAMES PRODUCTION AT--ETC(U)
JAN 82 D R CALDWELL
UNCLASSIFIED ETL-R032 NL

1 OF 1
AD A
110926

END
DATE
FILMED
03-82
DTIC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

2

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER R032	2. GOVT ACCESSION NO. AD-A110926	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A NAME INPUT STATION FOR AUTOMATED FOREIGN NAMES PRODUCTION AT THE UNITED STATES DEFENSE MAPPING AGENCY		5. TYPE OF REPORT & PERIOD COVERED Paper
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) DOUGLAS R. CALDWELL		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS (12) 17		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories ATTN: ETL-LO Ft. Belvoir, VA 22060		12. REPORT DATE 29 Jan 82
		13. NUMBER OF PAGES 15
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

geographic names
Foreign Place Names File
computer
diacritics
special symbols
fonts
Regional Diactritic Sets (REDS)
gazetteers
automated names production

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The United States Defense Mapping Agency is the federal organization charged with maintaining foreign names information. The Foreign Place Names File contains references to more than 3.5 million approved and 1.5 million variant names for foreign countries, undersea features, and extraterrestial features. The file is currently stored on index cards, because existing cumputer equipment has not been capable of displaying and processing the diacritics and special symbols characteristic of the extended Latin alphabet. In a joint effort with the US Army Engineer Topographic Laboratories and the Illinois Institute of Technology

403192

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BLOCK #20

and Research Institute, the United States Defense Mapping Agency is developing a prototype Names Input Station to digitally process foreign names data. The station can input, output, edit, and display diacritics and special symbols for the over one hundred romanized non-English languages required by the agency.

A
NAMES INPUT STATION
FOR
AUTOMATED FOREIGN NAMES PRODUCTION
AT THE
UNITED STATES DEFENSE MAPPING AGENCY

DOUGLAS R. CALDWELL
AUTOMATED CARTOGRAPHY BRANCH
US ARMY ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060

BIOGRAPHICAL SKETCH

Douglas R. Caldwell is the Project Engineer assigned to the development of the prototype Names Input Station. Mr. Caldwell received his BA Degree in Geography from Dartmouth College and MS Degree in Cartography from the University of Wisconsin. He is currently involved with a number of Mapping, Charting and Geodesy research projects at the US Army Engineer Topographic Laboratories.

ABSTRACT

The United States Defense Mapping Agency is the federal organization charged with maintaining foreign names information. The Foreign Place Names File contains references to more than 3.5 million approved and 1.5 million variant names for foreign countries, undersea features, and extraterrestrial features. The file is currently stored on index cards, because existing computer equipment has not been capable of displaying and processing the diacritics and special symbols characteristic of the extended Latin alphabet.

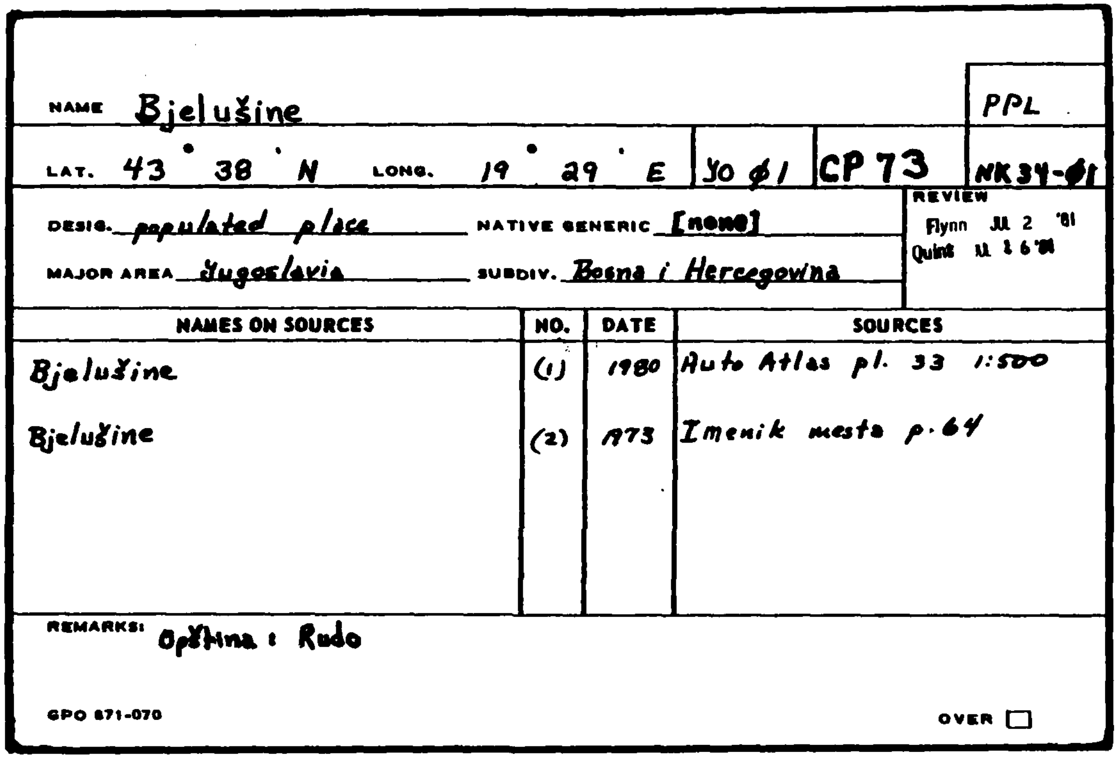
In a joint effort with the US Army Engineer Topographic Laboratories and the Illinois Institute of Technology and Research Institute, the United States Defense Mapping Agency is developing a prototype Names Input Station to digitally process foreign names data. The station can input, output, edit, and display diacritics and special symbols for the over one hundred romanized non-English languages required by the agency.

INTRODUCTION

The United States Defense Mapping Agency (DMA) is the federal agency in the United States responsible for maintaining names information on foreign places, undersea features, and extraterrestrial features. DMA collects and evaluates names data and works with the United States Board on Geographic Names (BGN) to "develop policies, principles, and procedures governing the use, spelling, and application of geographic names."[1] In addition, DMA maintains the BGN Foreign Place Names File, which is used to produce gazetteers, validate information for hydrographic, topographic, and aeronautical map and chart products, and as a resource for responding to inquiries from other government agencies and private businesses or individuals.

The cornerstone of the names production process, the Foreign Place Names File, consists of names, locations, descriptions, and source material history for geographic features. (See Figure 1) Stored on 4" X 6" index file cards, this massive file contains more than 3.5 million BGN approved names and over 1.5 million recorded variants. This present manual system is unable to keep pace with ever increasing demands for new and updated names information. Initial attempts to automate the system were hindered, because existing computer equipment could not adequately process and display the names data.

Geographic names information at DMA is unique because of its extended Latin character set. Foreign languages based on the Roman alphabet may contain diacritics and special symbols. French, for example, has acute (´) and



NAME Bjelušine

PPL

LAT. 43° 38' N LONG. 19° 29' E

YO ø1 CP 73 NK34-ø1

DESIG. populated place NATIVE GENERIC [none]

MAJOR AREA Yugoslavia SUBDIV. Bosna i Hercegovina

REVIEW
Flynn JUL 2 '81
Quinn JUL 16 '81

NAMES ON SOURCES	NO.	DATE	SOURCES
Bjelušine	(1)	1980	Auto Atlas pl. 33 1:500
Bjelušine	(2)	1973	Imenik mesta p. 64

REMARKS: Opština : Rudo

GPO 871-070

OVER □

Figure 1

Sample Index Card from the Foreign Place Names File grave (\`) accents, circumflexes (^), umlauts (¨), and cedillas (¸) for diacritics, and the o e ligature (œ) and apostrophe (ʾ) for special symbols. Also, non-Roman alphabets may be phonetically converted to Roman based forms using transliteration schemes. Thus, in Arabic, an آ becomes an ā. All told, DMA uses one hundred and sixteen languages, of which seventy-three have approved diacritics or special symbols. More languages use diacritics and special symbols, but since the transliteration schemes are not approved, they are not included in the system. Some of the languages, like Chichewa, spoken in Malawi, may contain a single diacritic, but most have more; in fact, Vietnamese contains fifty-four diacritic and letter combinations and thirteen special symbols. While phototypesetting has long permitted the printing of special symbols and diacritics for final copy, the problem for the names specialist has been to locally display the special symbols or characters with diacritics located in the correct position. Consider the problems in printing an "a" with an acute accent (á). The earliest names information was printed in uppercase letters with no diacritics, or (A). Next, lowercase letters could be displayed, but there were still no diacritics (a). DMA currently has a high speed printing system which can display some diacritics adjacent to their associated character (a ´). Ultimately, however, the goal has been to print the diacritic in its proper position relative to its associated character (á) and this can be done on the Names Input Station.

The United States Defense Mapping Agency tasked the US Army Engineer Topographic Laboratories (USAETL) to design and build a prototype Names Input Station for automated names work. Development is being accomplished under contract with the Illinois Institute of Technology and Research Institute (IITRI) and is scheduled for completion in early 1982.

NAMES INPUT STATION

The Names Input Station (NIS) was developed to input, output, edit, and display names data, DMA outlined a number of features necessary for handling foreign languages. First, the keyboard had to be designed to allow an operator to enter a diacritic or special symbol with a single keystroke. Second, the station had to display the special symbols and diacritics in the proper position relative to its associated uppercase or lowercase alpha character. Third, the data had to be stored in ASCII format, so the information could be processed by any computer using this format, whether or not that computer could display the special symbols or diacritics. Fourth, the station had to provide local hard copy of the data displayed, again with correctly positioned diacritics.

The Names Input Station developed by IITRI includes three hardware items: an ECD Smart ASCII intelligent terminal, a Per Sci floppy disk drive, and a Florida Data printer. The ECD Smart ASCII was selected for its flexibility and met all but the hard copy requirements. In addition to a standard terminal keyboard, the Smart ASCII has two outboard keypads which give it a total of one hundred and thirty-four keys. The additional keys are necessary for the placement of the special symbols and diacritics. A close-up of the left outboard keypad (See Figure 2) shows examples of the diacritics located over a representative character. To enter a character and diacritic, an operator would first strike the key with the character on the central standard keyboard, and then move to the outboard keypad and strike the key with the diacritic. The resulting character and

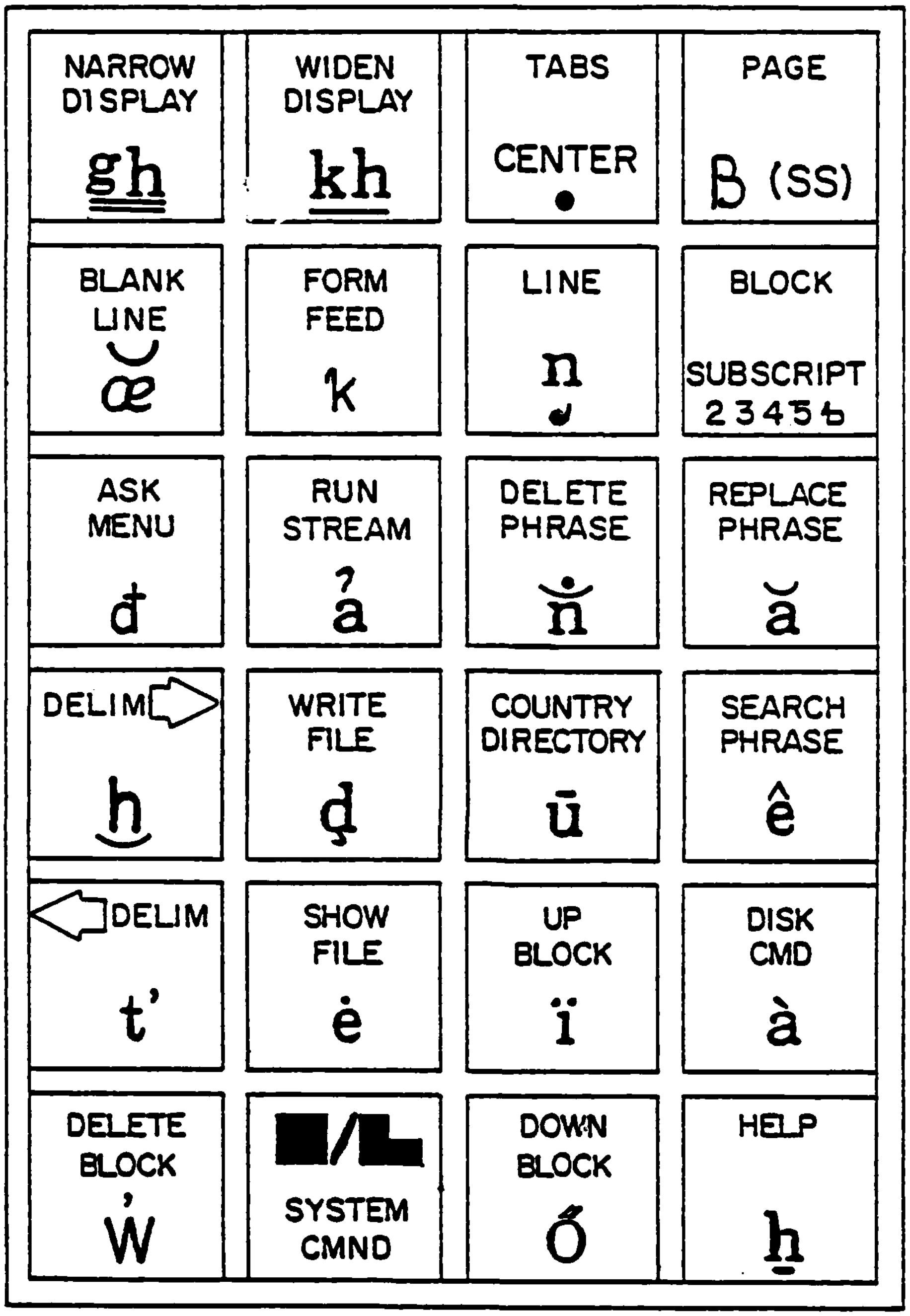


Figure 2

Left Outboard Keypad of Names Input Station

The diacritics are displayed with a sample character. For example, the circumflex (^) over the e could also be placed over an a, i, o, or u if the names specialist were entering French names data.

properly positioned diacritic will then appear on the cathode ray tube (CRT) display to verify the input. Hard copies of the screen data may next be printed on the ECD modified Florida Data Model BNY matrix printer.

NAMES INPUT STATION SOFTWARE

The Names Input Station software has been developed by the Illinois Institute of Technology Research Institute around the ECD program Translex, a powerful text editing and word processing package. IITRI first exploited the user definable fonts to create a series of Regional Diacritic Sets (REDS) containing DMA's required special symbols and diacritics. IITRI then utilized the Translex macro-programming facility and developed software routines to meet the names production needs.

The creation of diacritics and special symbols is central to the Names Input Station software. Each individual character and diacritic is defined by an 8 X 16 matrix of cells. (See Figure 3) These are edited and refined to meet the cosmetic requirements of DMA. The special symbols and diacritics are then assigned to keys on the outboard keypad. The keyboard may be redefined at any point to change or expand the system to accommodate revised or newly approved transliteration schemes. This flexibility is highly desirable in a research system.

The diacritics and special symbols are grouped into Regional Diacritic Sets. Each REDS contains the diacritics and special symbols for a number of languages within geographically related countries. There are fourteen REDS encompassing the one hundred and sixteen languages required by DMA. A particular language may be found in more than one REDS, and most REDS

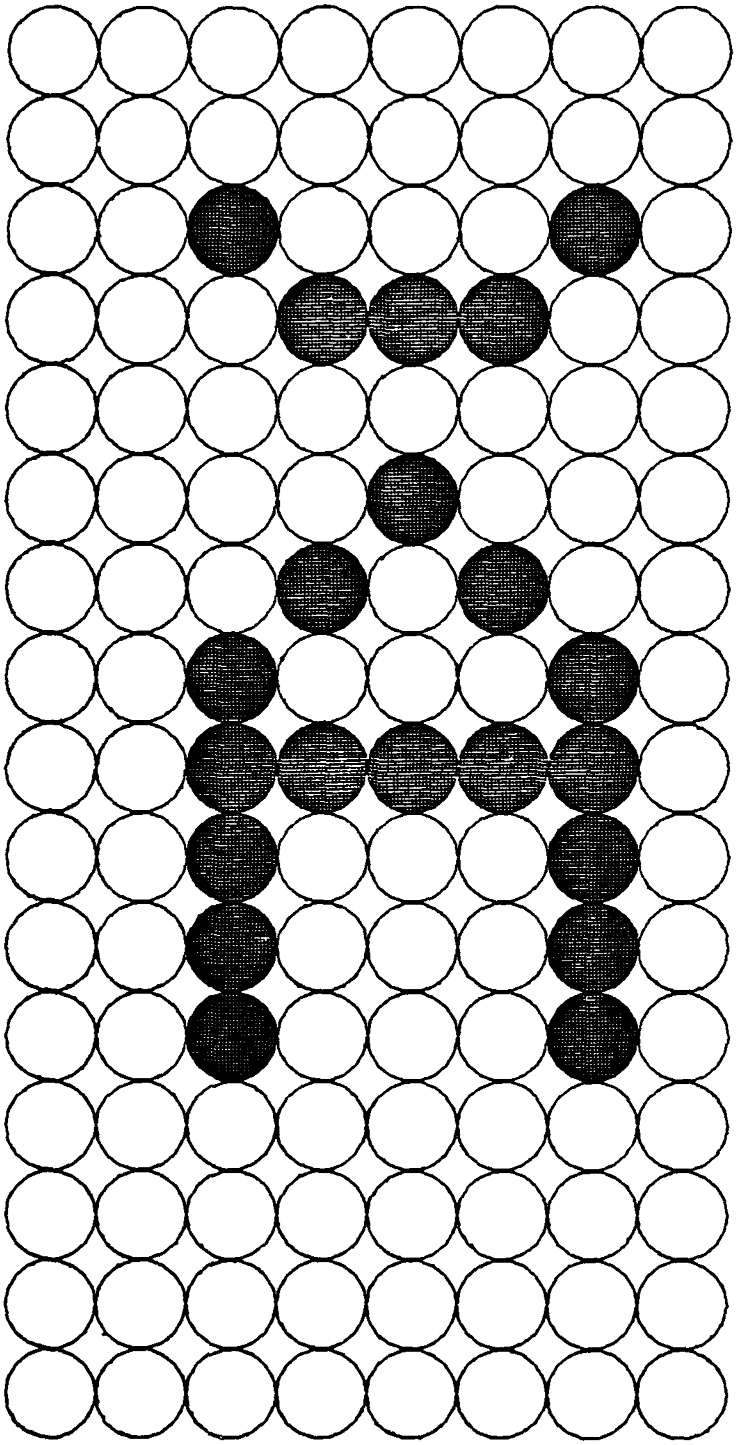


Figure 3

8 X 16 Matrix for Formation of a Single Character with Diacritic

The formation of an uppercase A with a breve (˘) is illustrated above. When the letter and diacritic are printed on the Florida Data printer, spaces between the cells on the matrix are filled in to produce a smoother image.

contain approximately seven languages. Thus, the REDS allow access to the diacritics and special symbols of the languages of a country, as well as neighboring countries. The only exceptions to the rule occur in the Soviet Union and Vietnam. The Soviet Union contains fifteen major languages and was broken down into three separate REDS. In Vietnam, the only major language was Vietnamese, but it contains such a large number of diacritics and special symbols that the REDS contains only a single language. All REDS are accessed by striking a REDS command key, typing the name of the country or language to be accessed, and striking the REDS command key once again. This loads the REDS so the appropriate characters appear on the CRT display and hard copy printouts. (See Figure 4)

Creating the individual characters and combining them into Regional Dacritic Sets preceded the development of software to meet DMA's names production needs. IITRI has also designed and implemented routines to augment gazetteer production, compile name change lists, and serve as a transliteration scratchpad.

DMA produces gazetteers for approximately one hundred and sixty-five countries, and each gazetteer contains information on the feature name, feature type (stream, mountain, etc.), latitude, longitude, UTM coordinate location, a numeric area code, and the Joint Operations Graphic map sheet number. The information is currently in both analog and digital form, with the digital data specially coded for phototypesetting operations and final gazetteer production. The Names Input Station will augment this process in two ways. First, the Florida Data printer will allow the names specialists to produce interim gazetteers on-site with no time delay. (See Figure 5)

North European Font

WHOLE FONT DISPLAY:

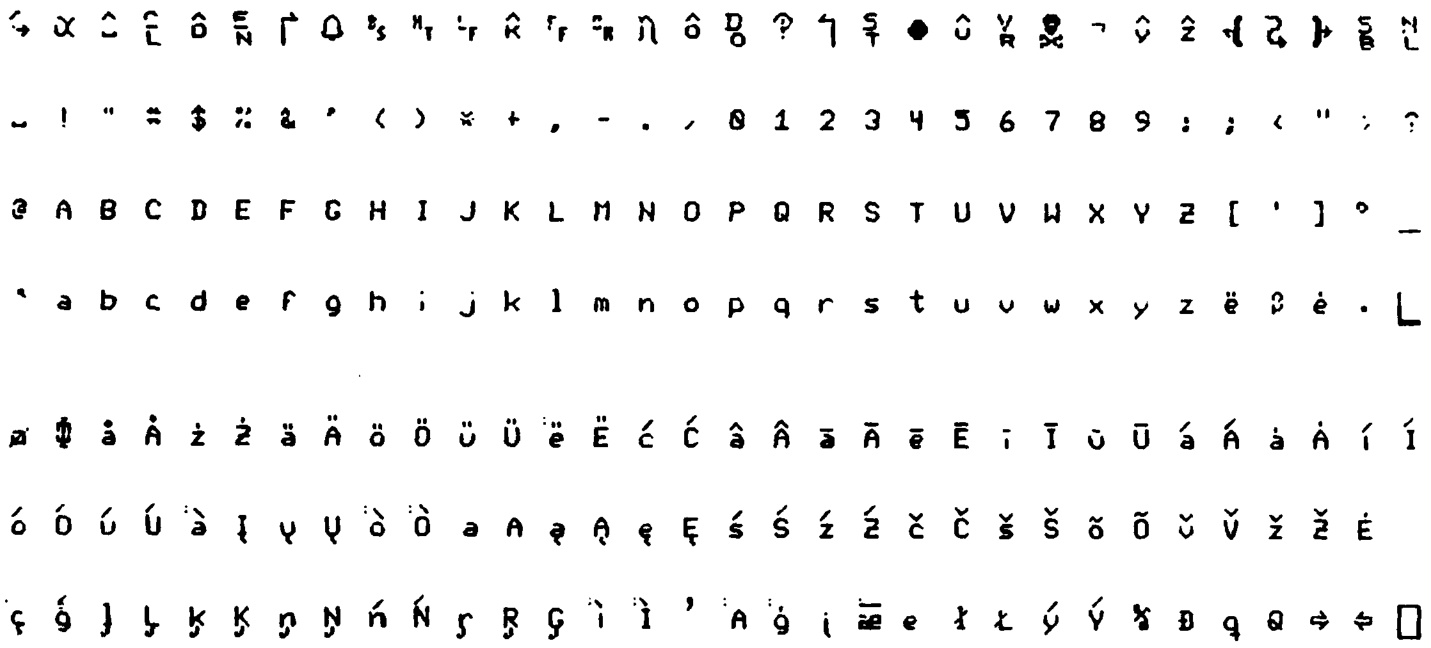
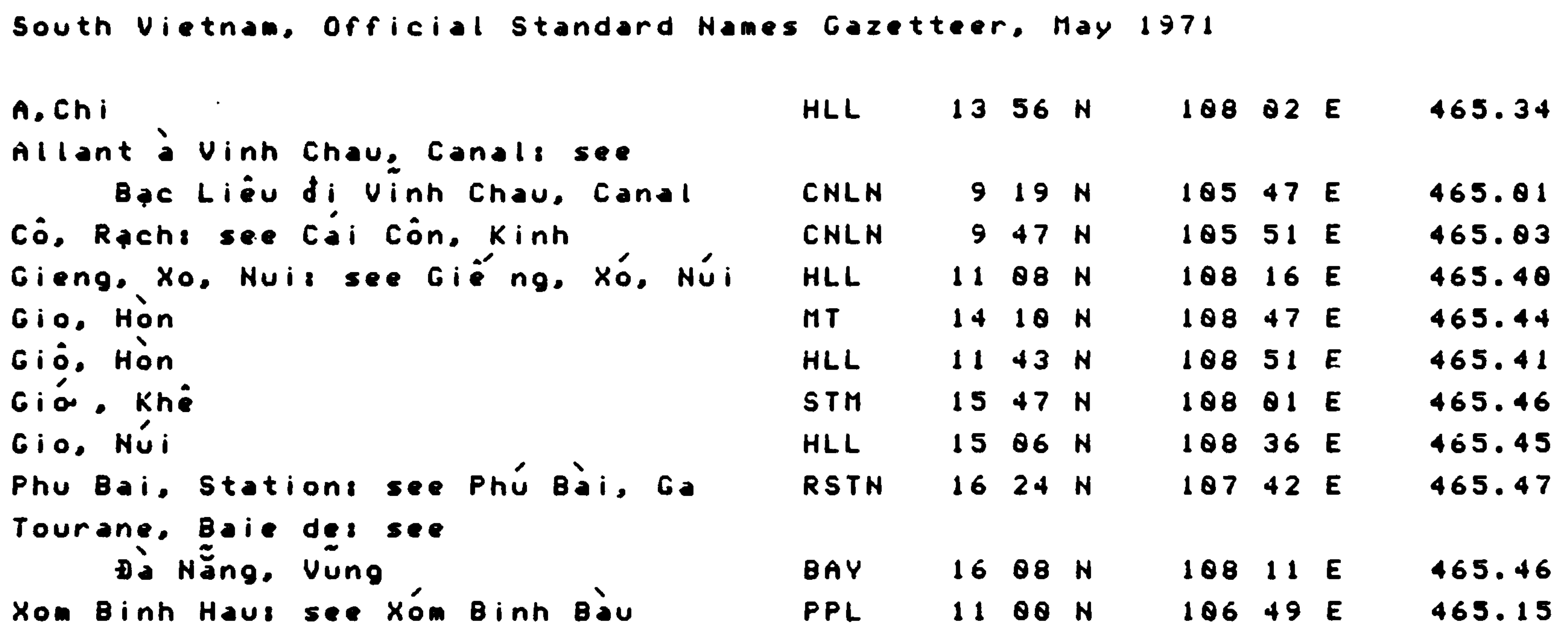

! " # $ % & ' ( ) * + , - . / 0 1 2 3 4 5 6 7 8 9 : ; < " > ?
@ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z ['] ° _
a b c d e f g h i j k l m n o p q r s t u v w x y z ë ė .

Figure 4

Sample Regional Diacritic Set (REDS)

The North European REDS contains all the DMA specified special symbols and diacritics for the languages of northern Europe. Languages or transliterated forms of languages covered include: Norwegian, Swedish, Danish, Finnish, Russian, Estonian, Latvian, Polish, German, Lappish, Lithuanian, Faroese, and Greenlandic.



South Vietnam, Official Standard Names Gazetteer, May 1971

A,Chi	HLL	13	56	N	108	02	E	465.34
Allant à Vinh Chau, Canal: see Bạc Liêu đi Vĩnh Chau, Canal	CNLN	9	19	N	105	47	E	465.01
Cô, Rạch: see Cái Côn, Kinh	CNLN	9	47	N	105	51	E	465.03
Gieng, Xo, Nui: see Giế ng, Xó, Núi	HLL	11	08	N	108	16	E	465.40
Gio, Hòn	MT	14	10	N	108	47	E	465.44
Giô, Hòn	HLL	11	43	N	108	51	E	465.41
Giớ , Khê	STM	15	47	N	108	01	E	465.46
Gio, Núi	HLL	15	06	N	108	36	E	465.45
Phu Bai, Station: see Phú Bài, Ga	RSTN	16	24	N	107	42	E	465.47
Tourane, Baie de: see Đà Nẵng, Vũng	BAY	16	08	N	108	11	E	465.46
Xom Binh Hau: see Xóm Binh Bàu	PPL	11	00	N	106	49	E	465.15

Figure 5

Sample Interim Gazetteer

The Names Input Station can be used to produce interim gazetteers. This section of a Vietnamese gazetteer is printed in the old format. Newer gazetteers include the UTM coordinates and Joint Operations Graphic sheet number. An interim gazetteer does not have the printed quality of a phototypeset product, but can be rapidly produced locally by a names analyst.

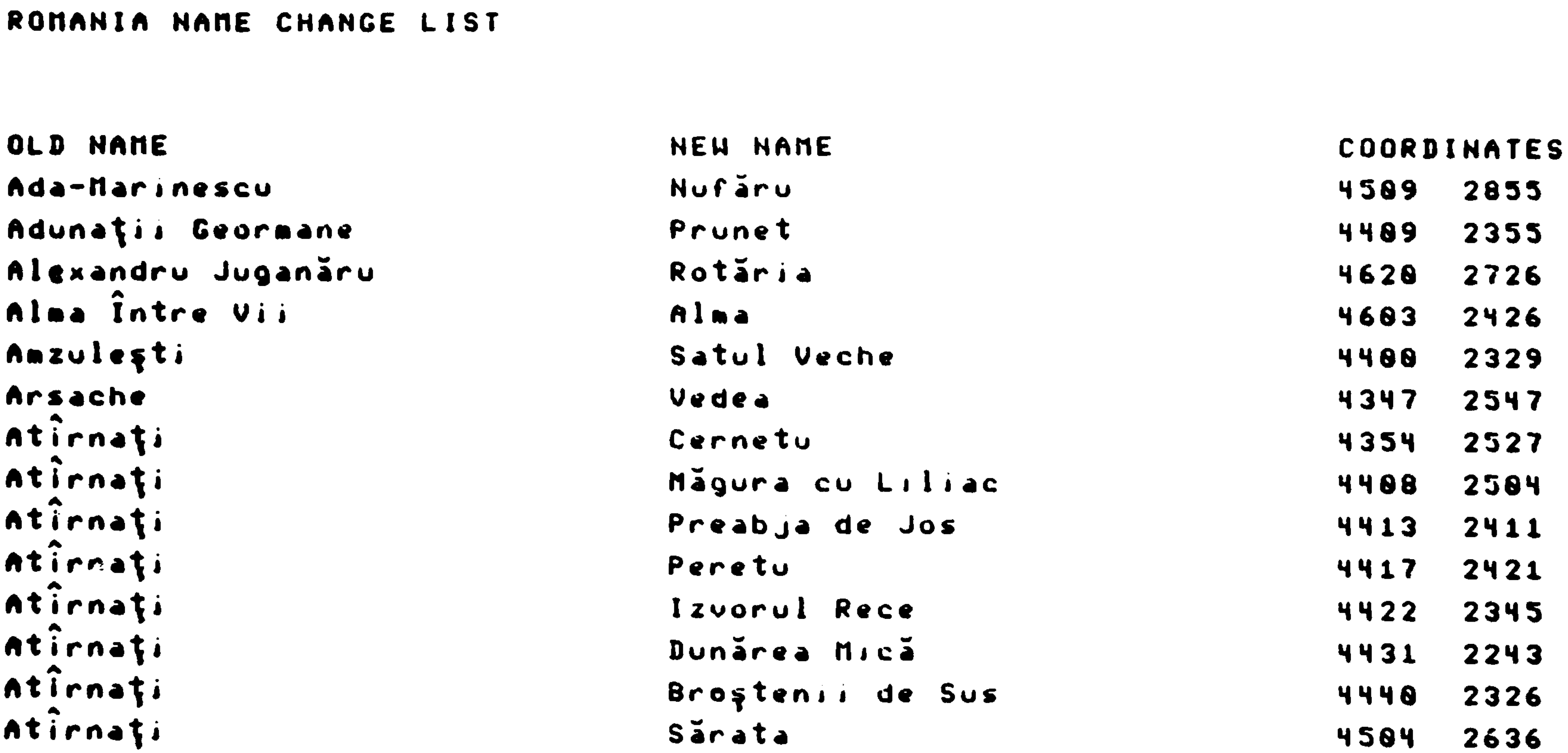
Second, the printer will provide local hard copy for data verification when the gazetteer information is to be transferred to the phototypesetter. Supplemental information used to support gazetteer production may be compiled using the name change list or transliteration scratchpad software.

The name change list (See Figure 6) contains three components: the old name, new name, and coordinate location. This information was formerly entered with an IBM Selectric typewriter with the diacritics later added by hand in a separate operation. With the use of the Names Input Station, this can be done with a single typing step.

The transliteration scratchpad function allows a names specialist to enter and store text in any of the available languages. As an example, a linguist might wish to record supplementary information on a particular transliteration scheme. (See Figure 7) The data could be stored in a file on disk and accessed for reference purposes by other linguists.

FUTURE ROLE OF THE NAMES INPUT STATION AT DMA

The Names Input Station fills both short term and long term needs of the Scientific Data Department at DMA. In the short term, the station addresses production needs by tying into the existing work flow. In its long term and more important role, the Names Input Station will serve as a research tool for the development of specifications for future automated names production, in particular, the creation of a digital Foreign Place Names File. The Names Input Station will be used to develop requirements for future data entry systems, evaluate the problem of standardization of diacritics and special symbols in an automated environment, define names



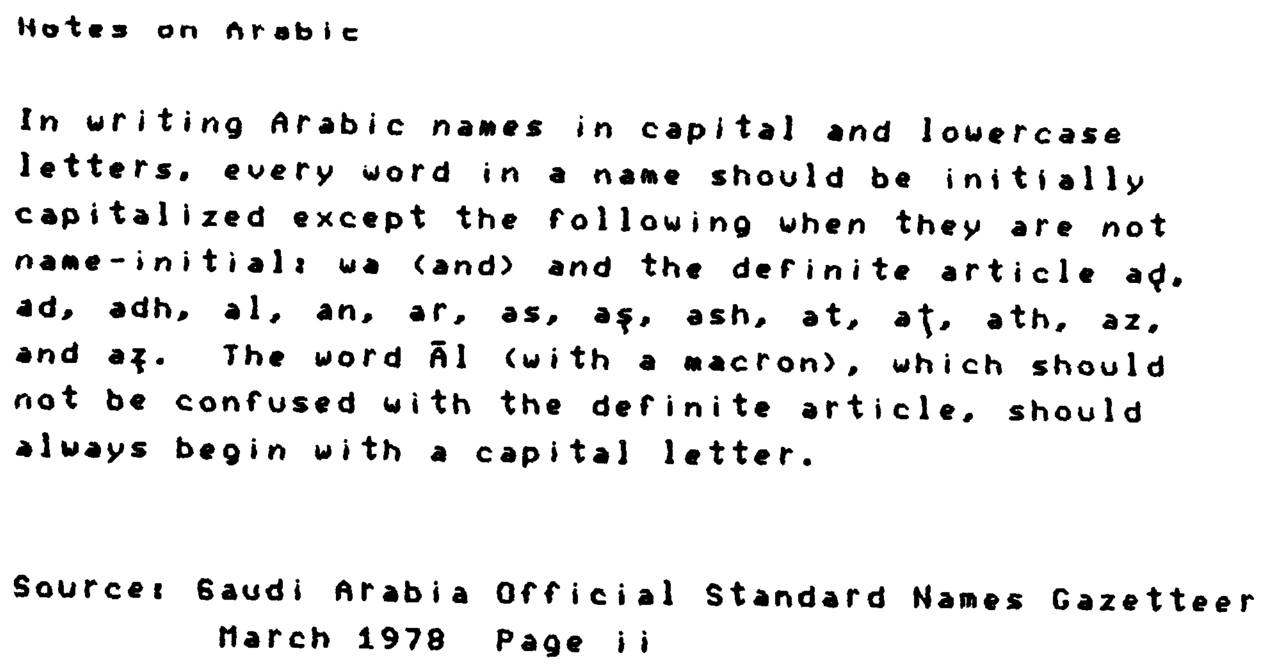
ROMANIA NAME CHANGE LIST

OLD NAME	NEW NAME	COORDINATES
Ada-Marinescu	Nufăru	4509 2855
Adunaţii Geormane	Prunet	4409 2355
Alexandru Juganăru	Rotăria	4620 2726
Alma Între Vii	Alma	4603 2426
Amzuleşti	Satul Veche	4400 2329
Arsache	Vedea	4347 2547
Atîrnaţi	Cernetu	4354 2527
Atîrnaţi	Măgura cu Liliac	4408 2504
Atîrnaţi	Preabja de Jos	4413 2411
Atîrnaţi	Peretu	4417 2421
Atîrnaţi	Izvorul Rece	4422 2345
Atîrnaţi	Dunărea Mică	4431 2243
Atîrnaţi	Broşteni de Sus	4440 2326
Atîrnaţi	Sărata	4504 2636

Figure 6

Sample Name Change List

Name change lists are compiled to keep the Foreign Place Names File up-to-date.



Notes on Arabic

In writing Arabic names in capital and lowercase letters, every word in a name should be initially capitalized except the following when they are not name-initial: wa (and) and the definite article aḍ, ad, adh, al, an, ar, as, aṣ, ash, at, aṭ, ath, az, and aẓ. The word Āl (with a macron), which should not be confused with the definite article, should always begin with a capital letter.

Source: Saudi Arabia Official Standard Names Gazetteer
March 1978 Page ii

Figure 7

Sample Transliteration Scratchpad

The transliteration scratchpad allows a names analyst to create, process, and store, transliterated text.

data formats, and analyze networking with other government and non-government computer systems.

Footnotes

[1] "United States Board on Geographic Names," pamphlet BGN/4/1980, p. 1.

Acknowledgement

The author would like to thank the following individuals for their assistance in preparing this paper: Robert F. Augustine (DMA), Charles Heyda (DMA), James Moore (IITRI), Diane Strife (IITRI), and Larry Cook (USAETL).

DATE
ILME